Basic Statistics

with Google Sheets

MORNING CLASS: 8.00 - 9.30 AM

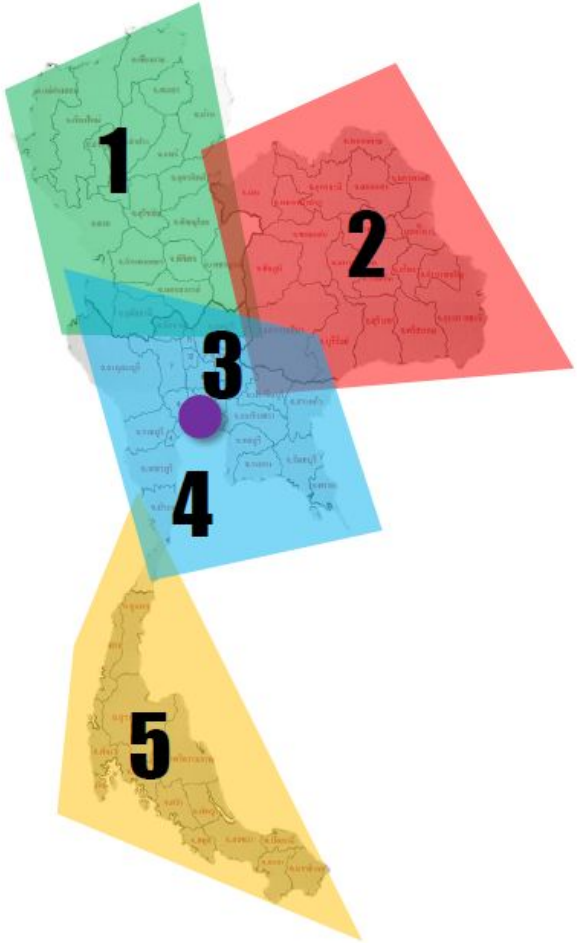
October 17, 2020

Today Agenda

1. Intro to Statistics
2. Hands-on exercises (Google Sheets)
    a. IMDb dataset
    b. Normal distribution
    c. Plotting distribution
    d. Measures of central tendency
    e. Measures of spread
    f. Measures of position
    g. Measures of relationship - correlation
    h. Measures of relationship - crosstabs
    i. Quick AB Test

Research Design

Random / Representative


1
2
3
4
5

2.4 million responded

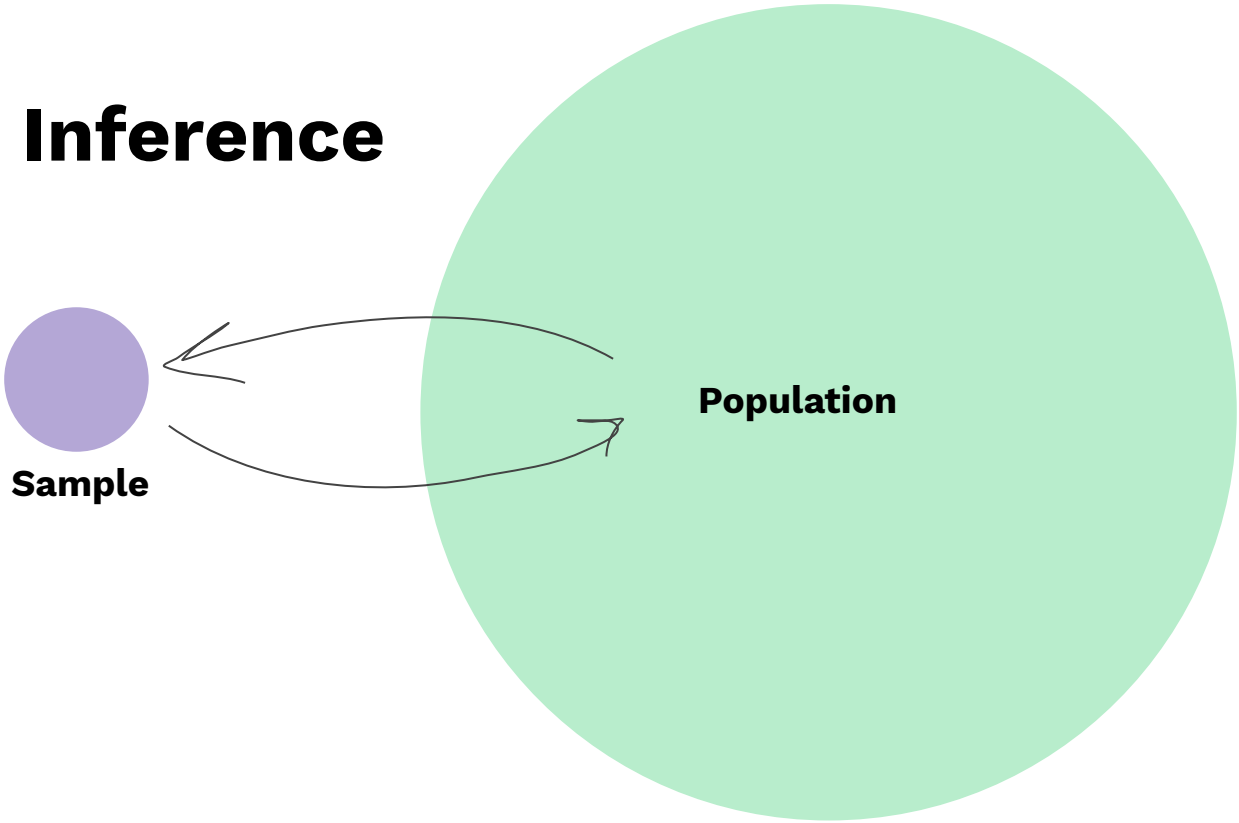
1936 predicted who was going to win the election

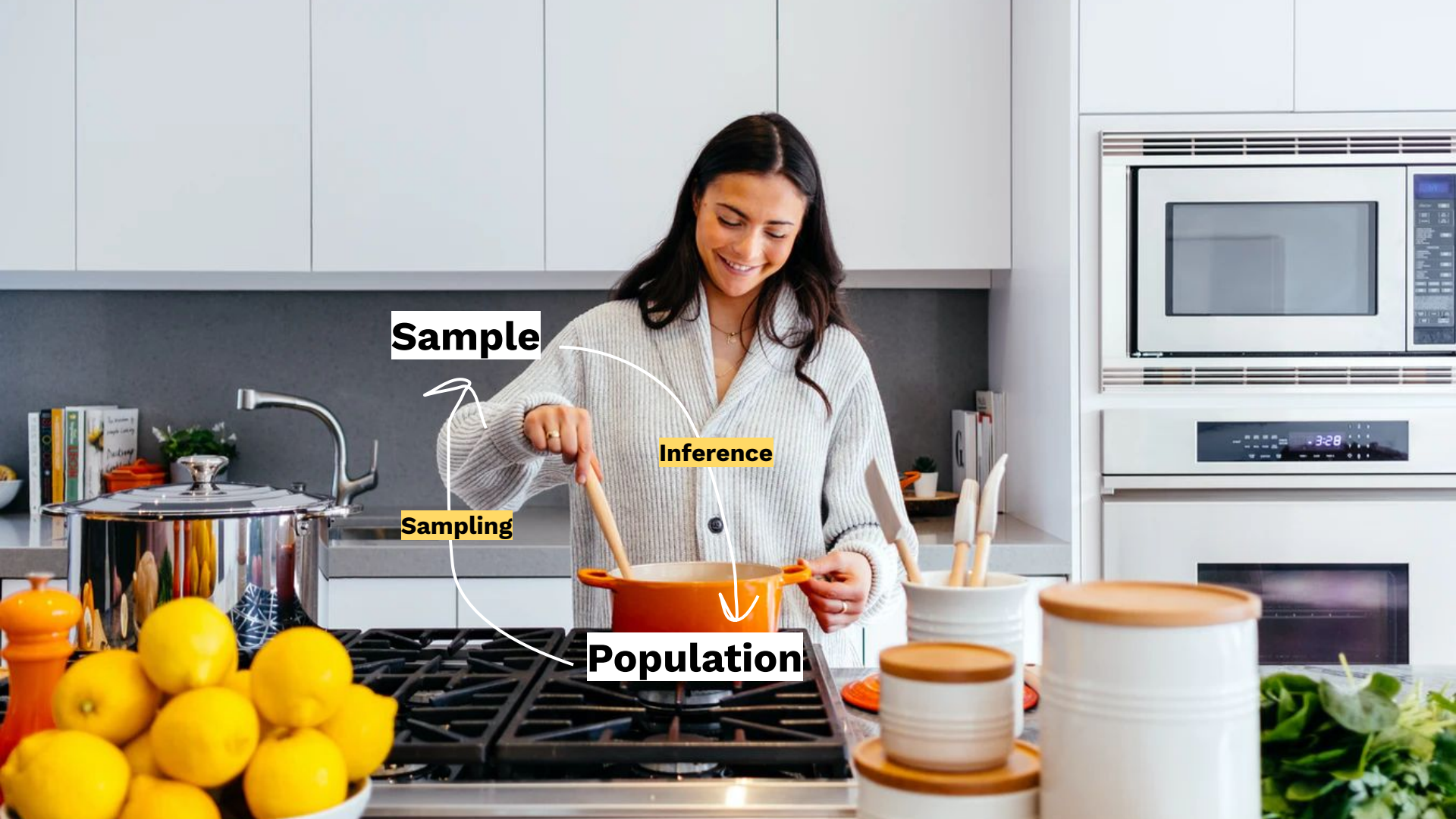
Alfred Landon vs. Franklin D. Roosevelt

https://en.wikipedia.org/wiki/The_Literary_Digest#Presidential_poll

Quality of our sample is of paramount importance

Inference




Sample
Inference
Sampling
Population

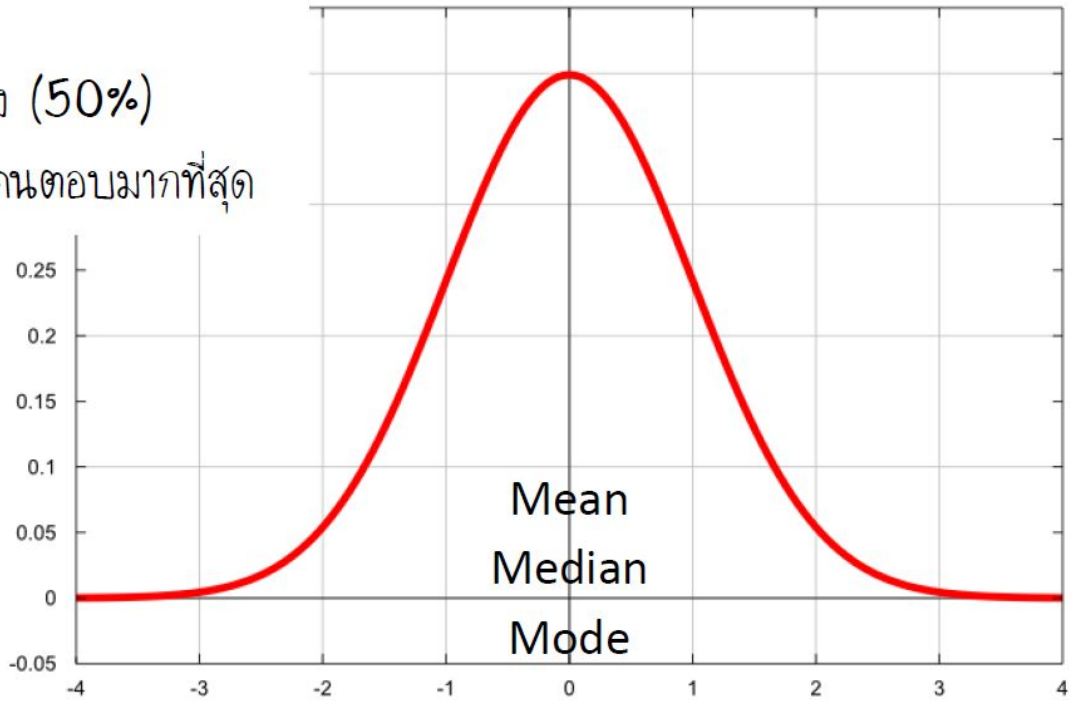
/The Descriptive Statistics_

Central Tendency

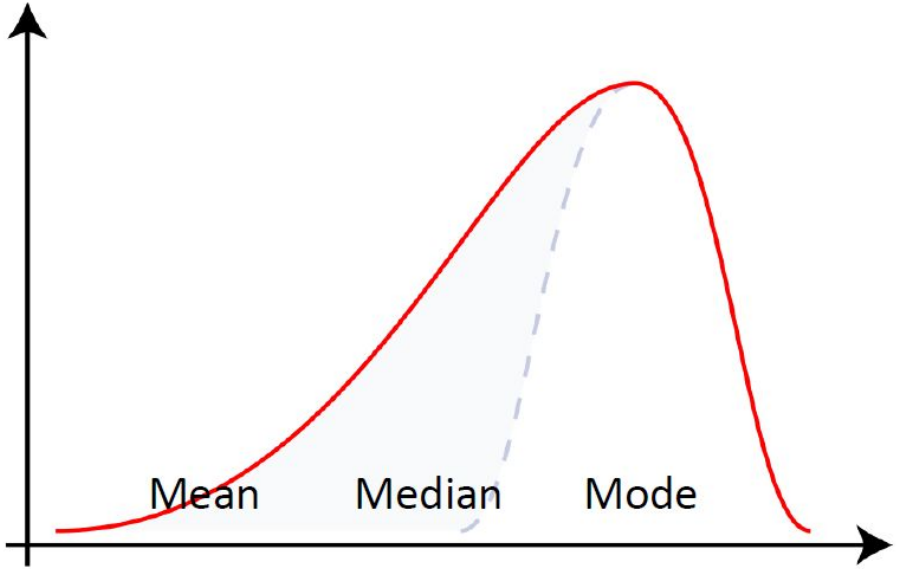
Mean = ค่าเฉลี่ย

Median = ค่ากลาง (50%)

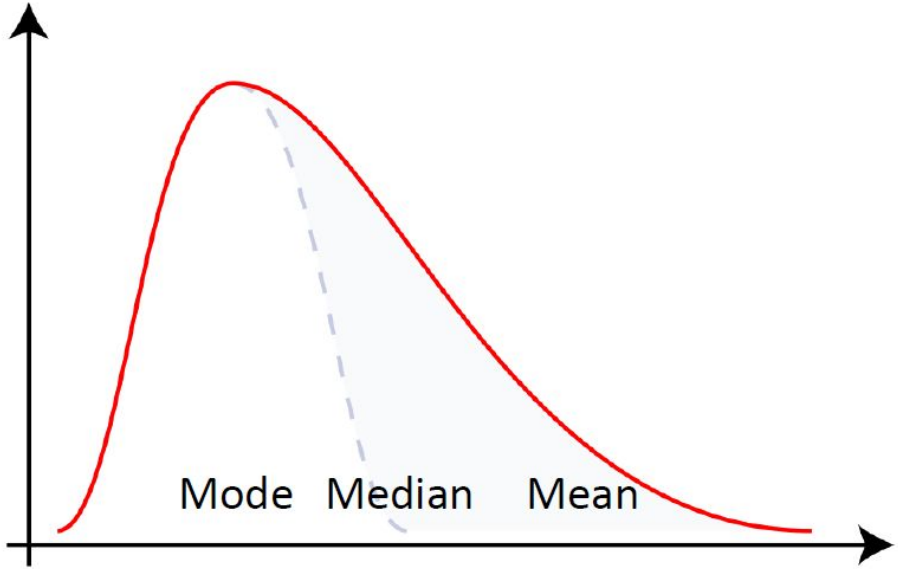
Mode = คำตอบที่คนตอบมากที่สุด



Skewed Distribution



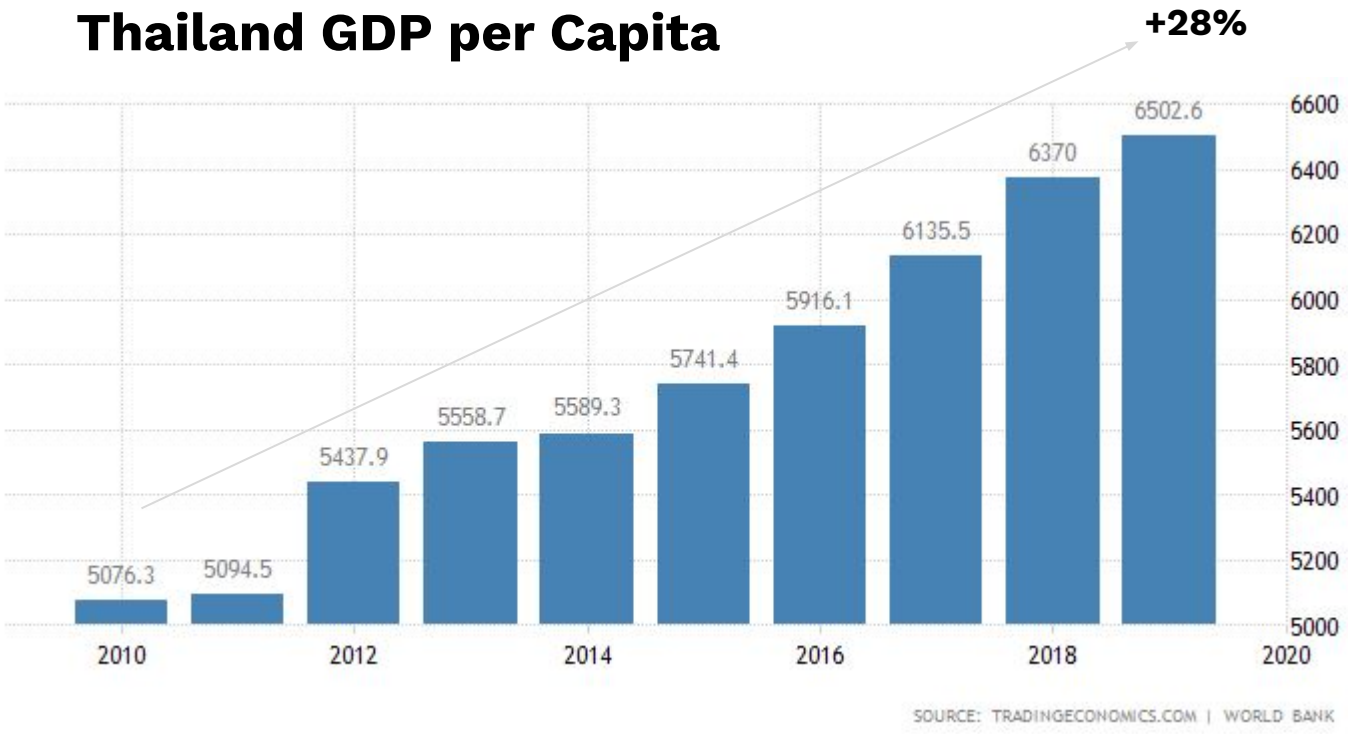
Negative Skew



Positive Skew

When to use **mean** or **median**
depends on **the distribution**

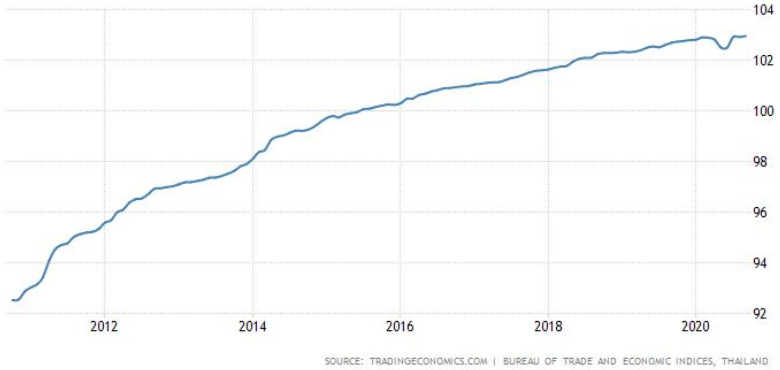
Thailand GDP per Capita



https://tradingeconomics.com/thailand/gdp-per-capita

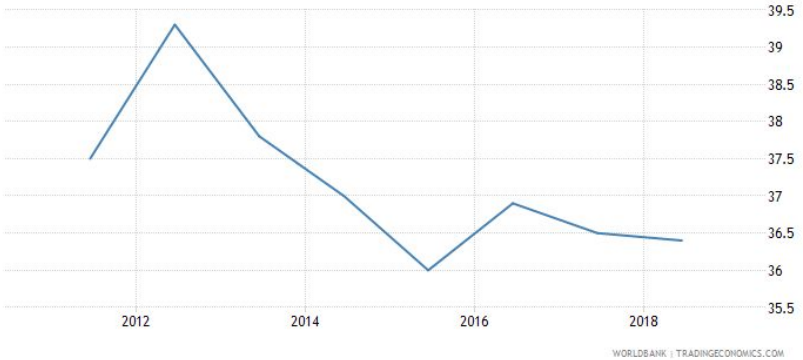
GDP = Price x Quantity

Core Consumer Prices

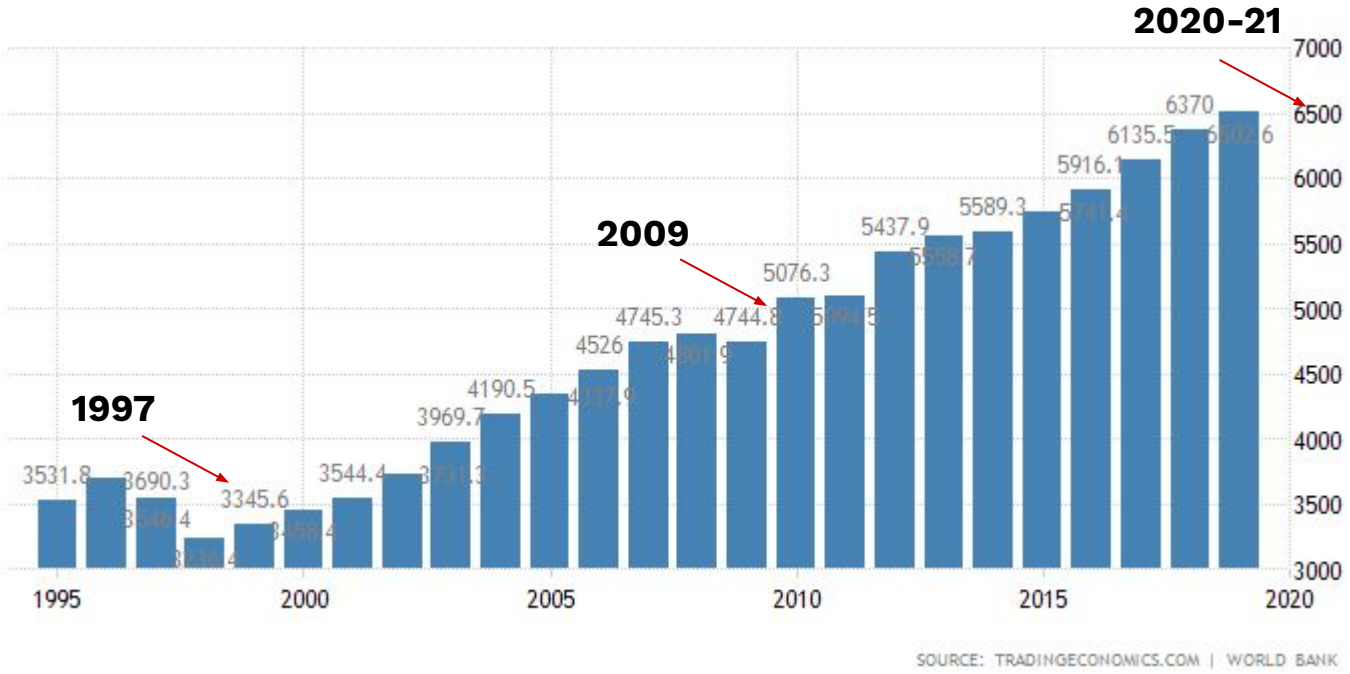


https://tradingeconomics.com/thailand/core-consumer-prices

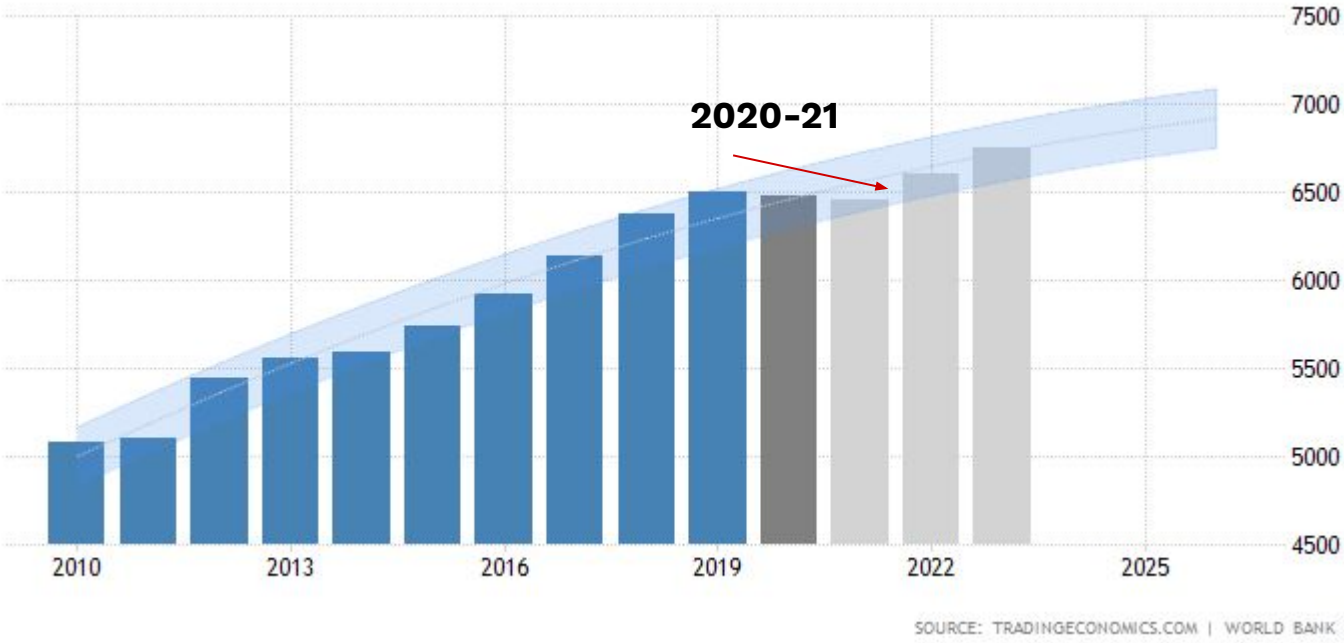
Gini Index



https://tradingeconomics.com/thailand/gini-index-wb-data.html

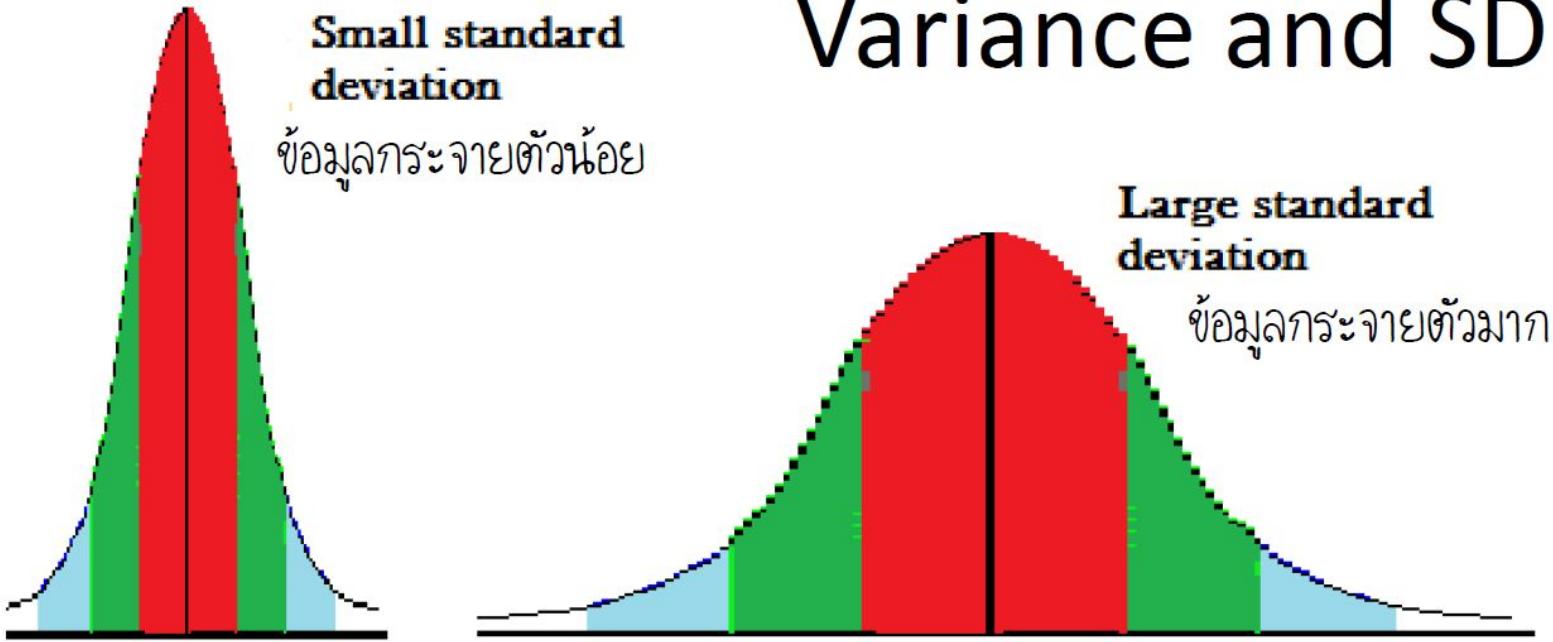


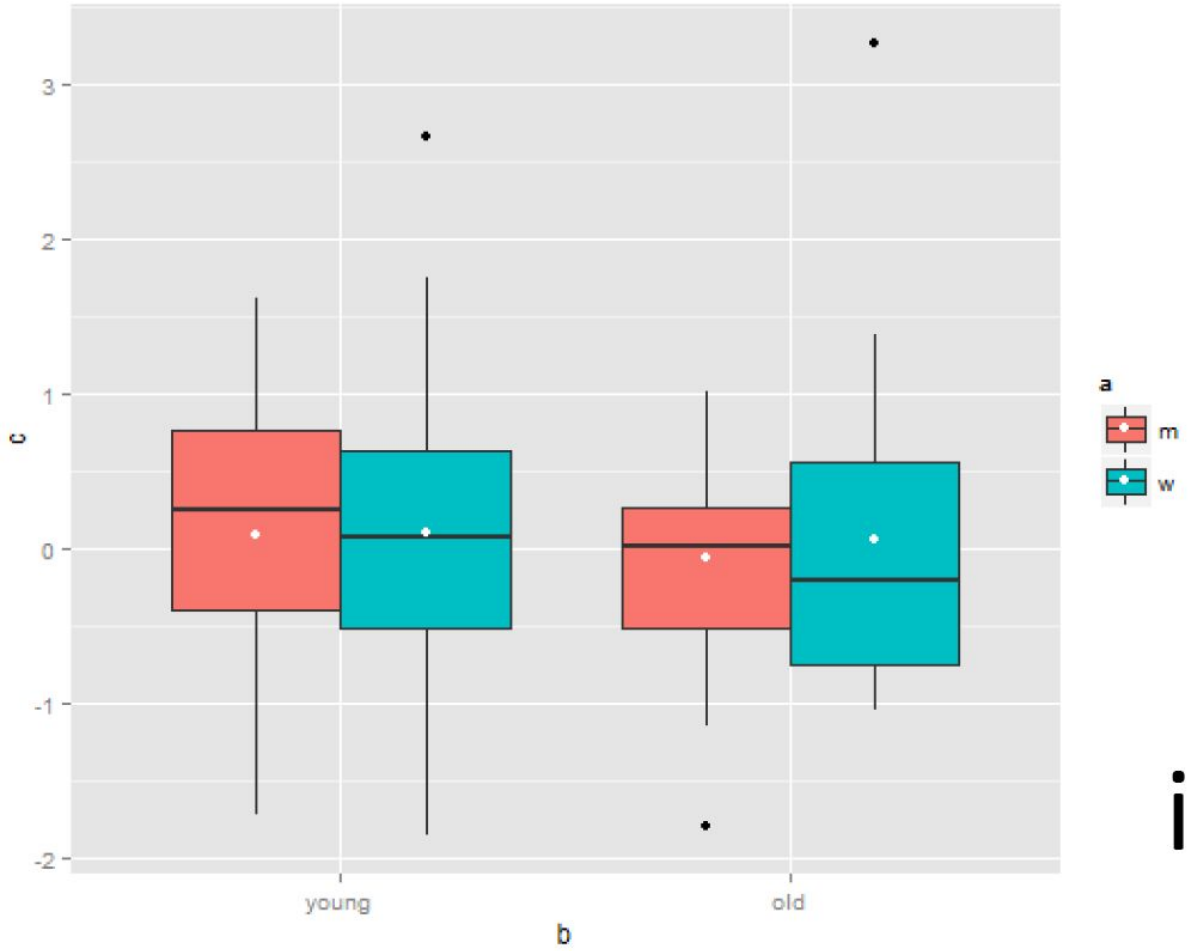
https://tradingeconomics.com/thailand/gdp-per-capita


2020-21
7500
7000
6500
6000
5500
5000
4500
2010
2013
2016
2019
2022
2025
SOURCE: TRADINGECONOMICS.COM | WORLD BANK

Variability

Variance and SD



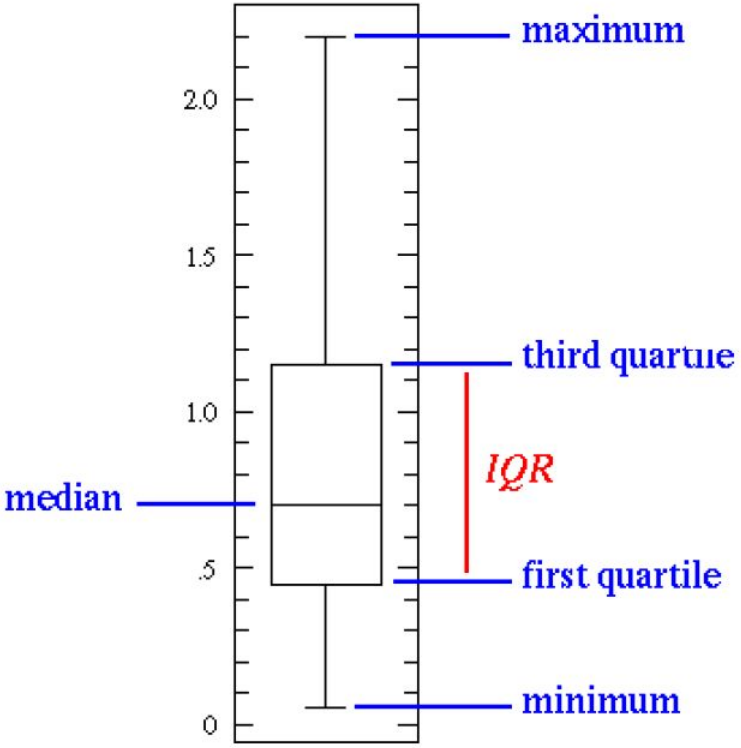


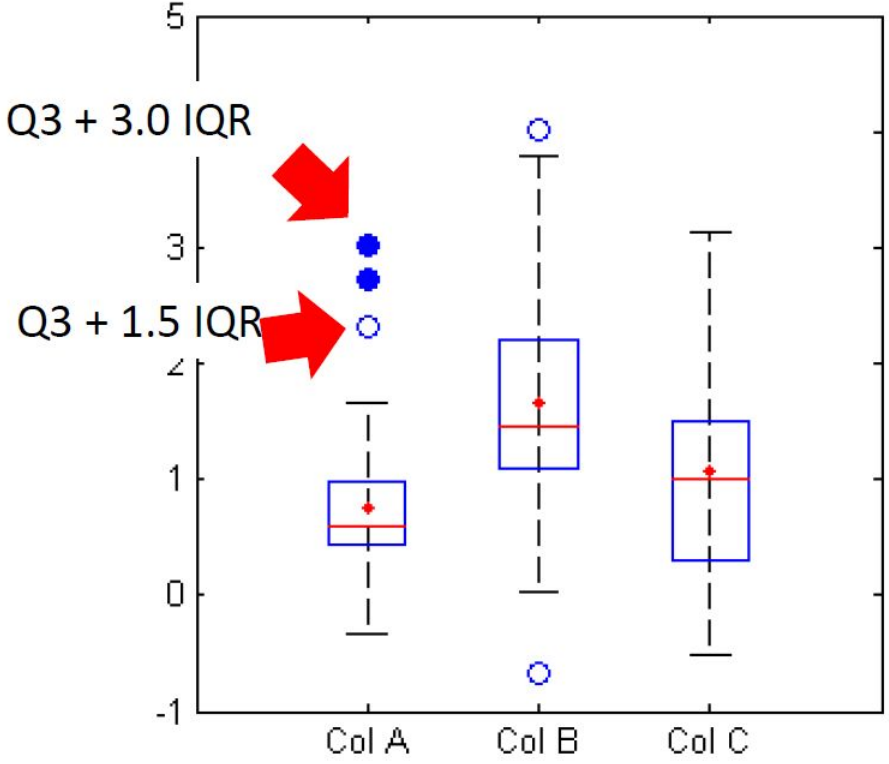
Position

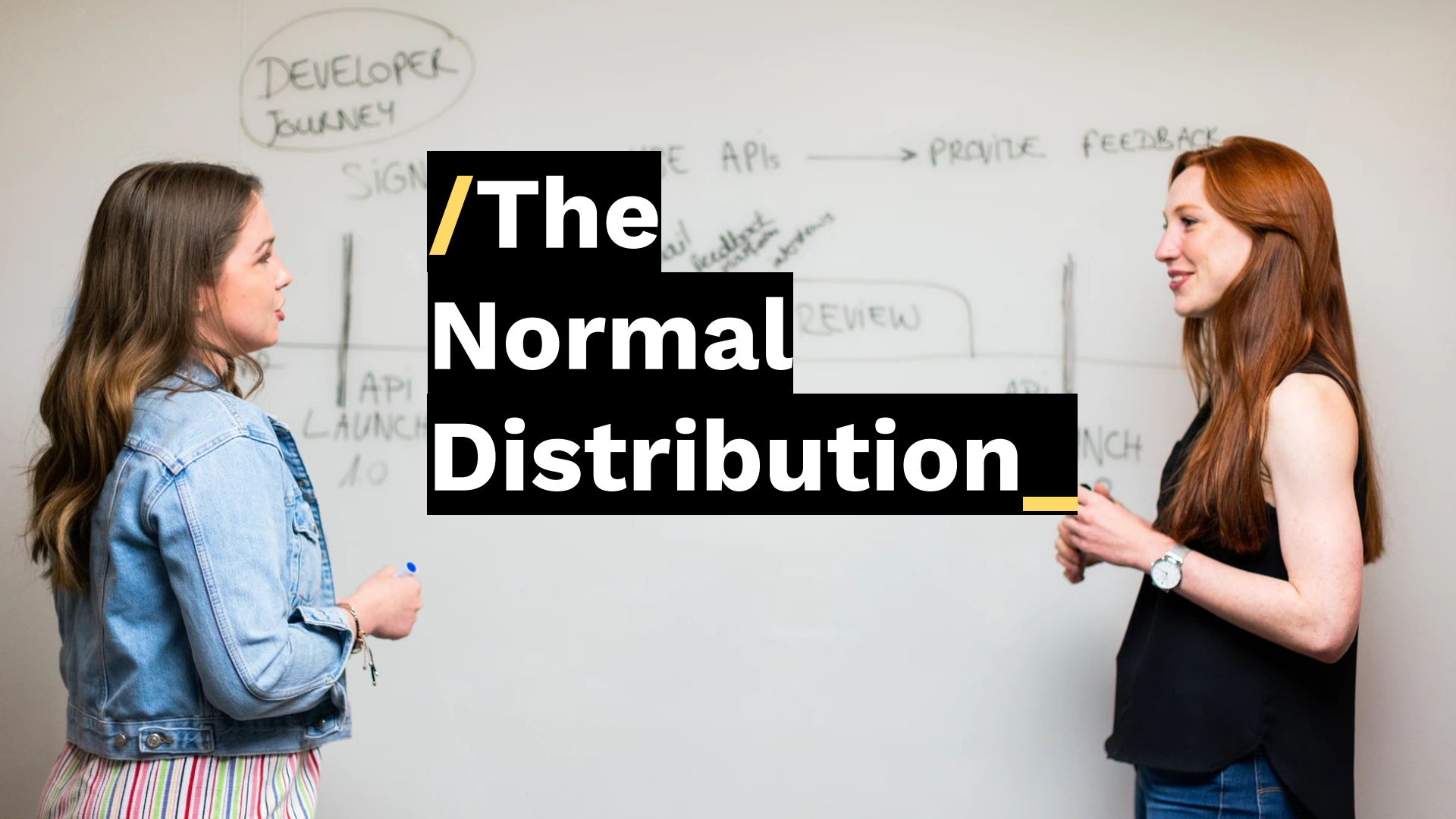
Boxplot

is good chart

Outliers

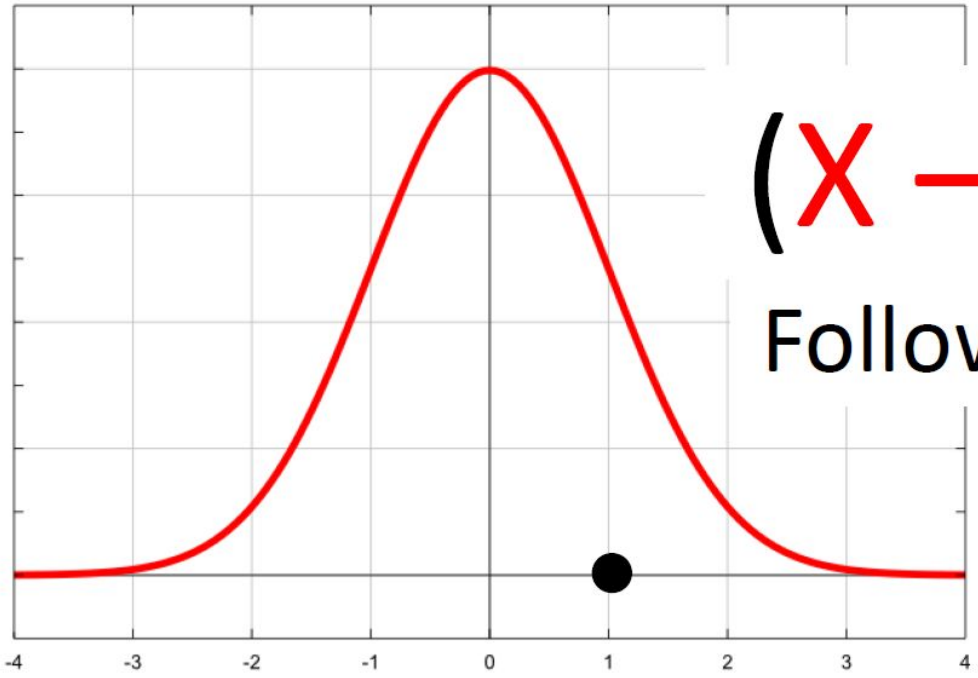

maximum
third quartile
IQR
median
first quartile
minimum
2.0
1.5
1.0
.5
0


Q3 + 3.0 IQR
Q3 + 1.5 IQR
5
3
1
0
-1
Col A
Col B
Col C



/The Normal Distribution_

Z-Score



(X – Mean) / SD

Follow normal distribution

Z 90% = 1.645

Z 95% = 1.960

Z 99% = 2.575

คนที่	ส่วนสูง	Z
1	163	-0.94
2	164	-0.78
3	180	1.63
4	168	-0.18
5	174	0.72
6	173	0.57
7	161	-1.24
8	160	-1.39
9	172	0.42
10	160	-1.39
11	180	1.63
12	175	0.87
13	164	-0.78
14	164	-0.78
15	178	1.33
16	173	0.57
17	175	0.87
18	168	-0.18
19	163	-0.94
20	169	-0.03

Z can be computed by statistical software

Mean	**169.20**
SD	**6.63**

ความหมายของค่า Z คือ ความสูงของคนนี้ (178 cm) สูงกว่าค่าเฉลี่ย (mean) คนไทย อยู่ 1.33 * SD

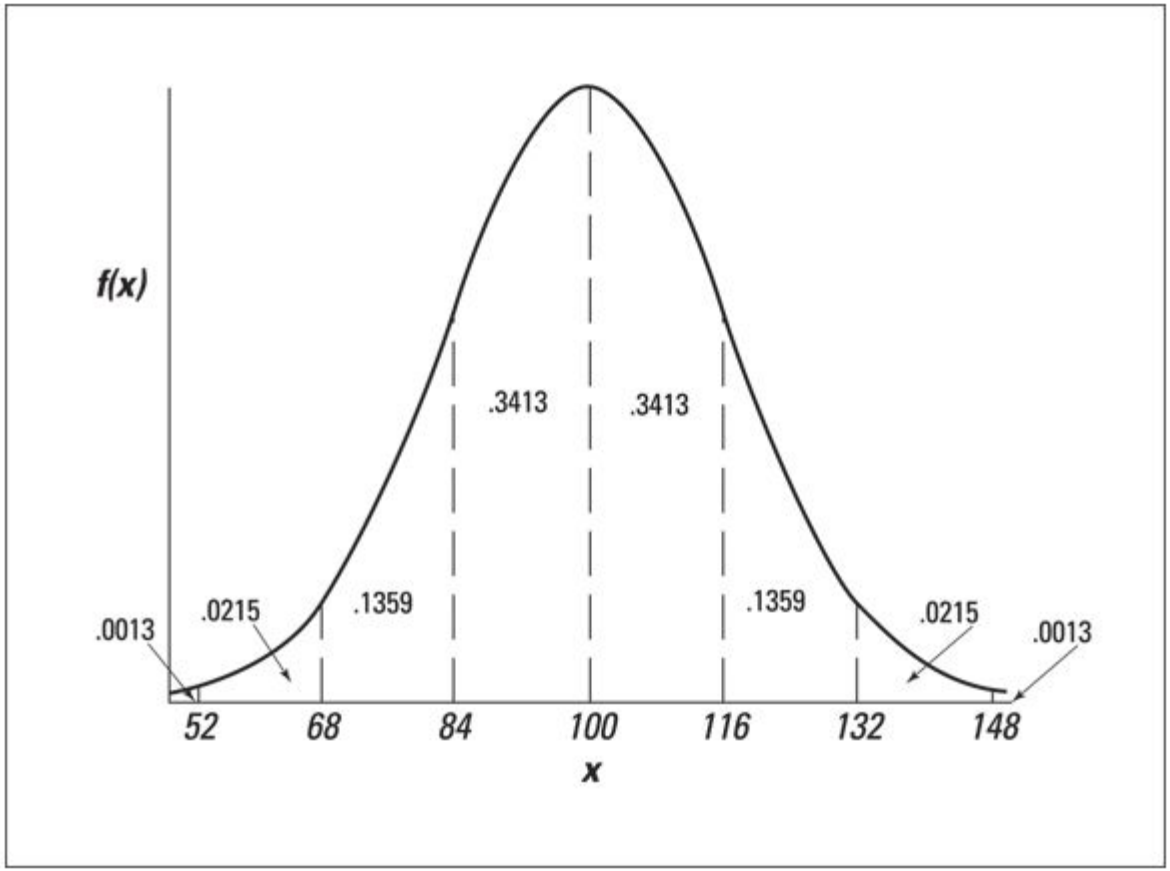
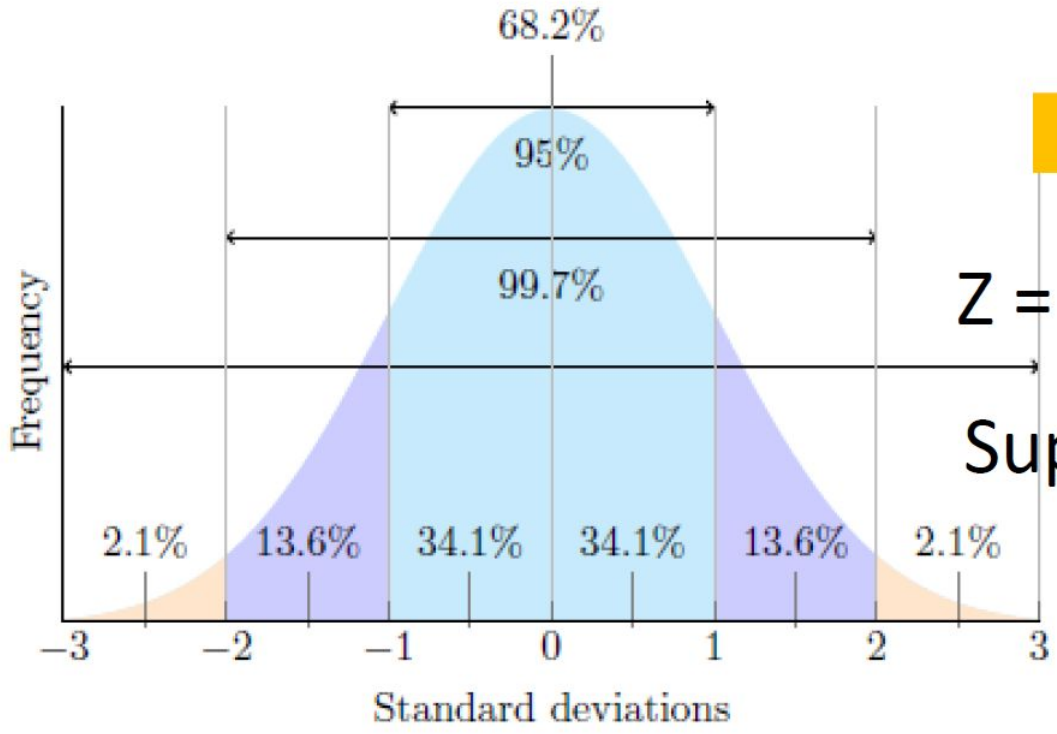


FIGURE 8-2: The normal distribution of IQ, divided into standard deviations.

Normal Distribution



Mean

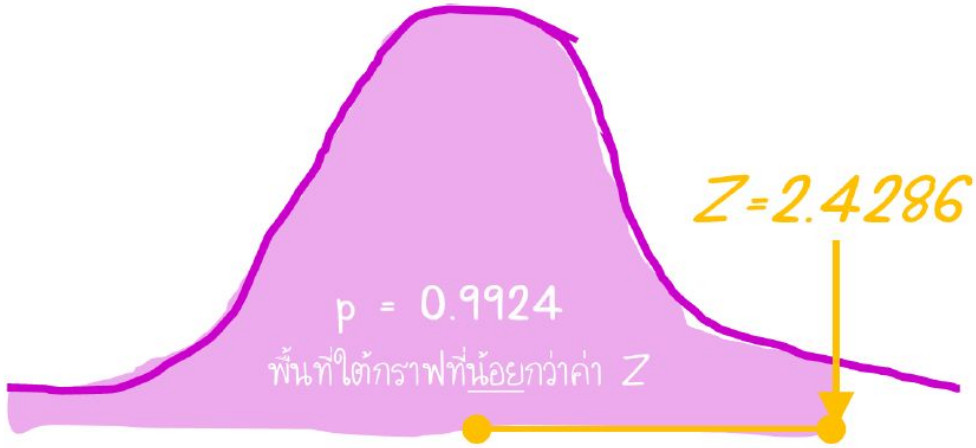
SD

$Z = N \sim (0, 1)$

Super Cool!

Z-Score of Test score

	A	B
1	**X**	72
2	**MEAN**	55
3	**SD**	7
4	**Cumulative**	TRUE
5		
6	**prob**	0.9924
7	**1-p**	0.0076




```
= NORM.DIST(72,55,7,TRUE)
```

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\mu =$ Mean

$\sigma =$ Standard Deviation

$\pi \approx 3.14159\cdots$

$e \approx 2.71828\cdots$

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Z

0

1

1

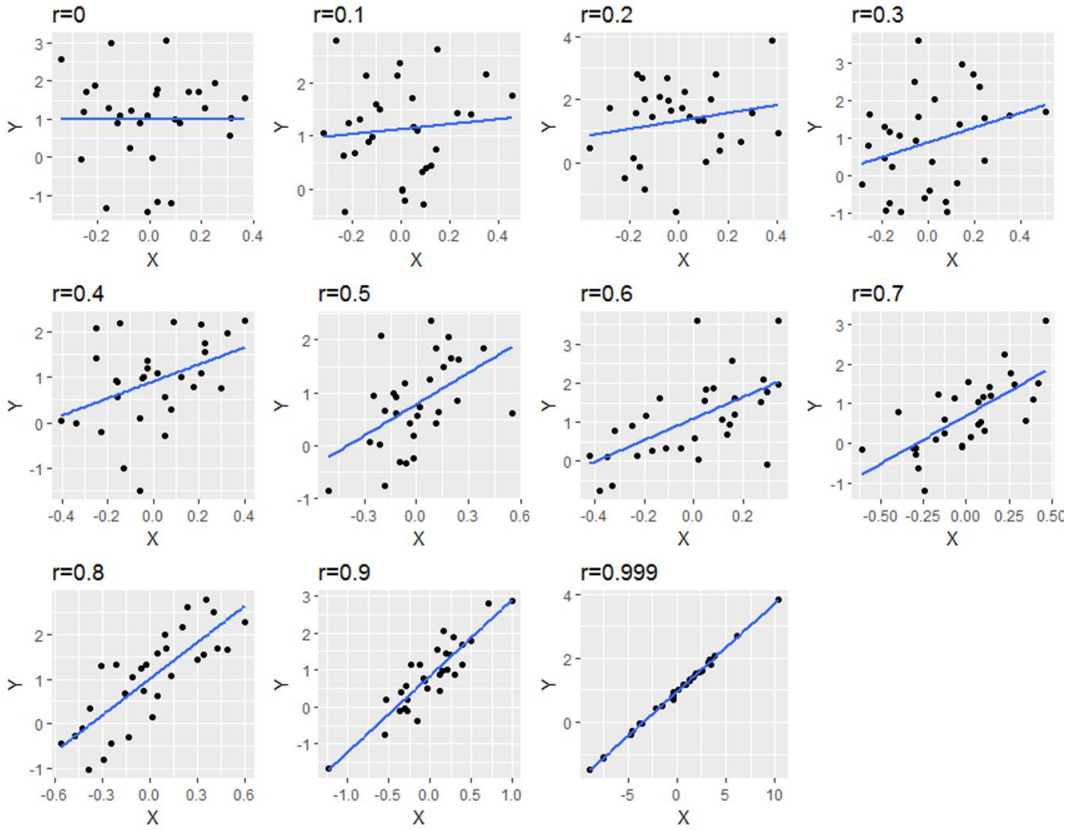
$\mu =$ Mean

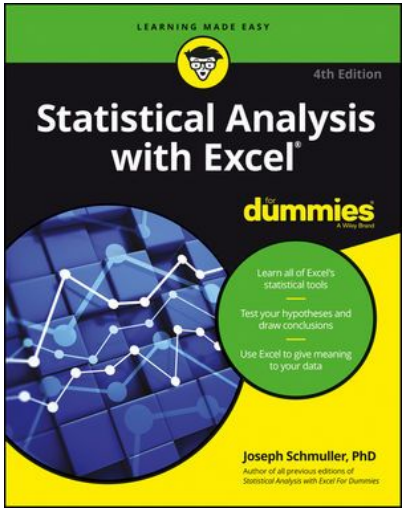
$\sigma =$ Standard Deviation

$\pi \approx 3.14159\cdots$

$e \approx 2.71828\cdots$

/The Measures of Relationship_


r=0
r=0.1
r=0.2
r=0.3
r=0.4
r=0.5
r=0.6
r=0.7
r=0.8
r=0.9
r=0.999
X
Y



Book Reference

Statistical Analysis with Excel for dummies (2016)

Basic Statistics

with Google Sheets

MORNING CLASS: 8.00 - 9.30 AM

SCHOOL OF FULL STACK
HTTPS://DATAROCKIE.COM

October 17, 2020